NAVAL HEALTH RESEARCH CENTER

A COMPARISON OF THE EFFECTS OF SLEEP DEPRIVATION ON SYNTHETIC WORK PERFORMANCE AND A CONVENTIONAL PERFORMANCE ASSESSMENT BATTERY

T. F. Elsmore
F. W. Hegge
P. Naitoh
T. Kelly
K. Schlangen
S. Gomez

Report No. 95-6

NAVAL HEALTH RESEARCH CENTER
P. O. BOX 85122
SAN DIEGO, CALIFORNIA 92186 – 5122

NAVAL MEDICAL RESEARCH AND DEVELOPMENT COMMAND
BETHESDA, MARYLAND

A Comparison of the Effects of Sleep Deprivation on Synthetic Work Performance and A Conventional Performance Assessment Battery[1]

Timothy F. Elsmore[2,3]
Frederick W. Hegge[2]
Paul Naitoh[4]
Tamsin Kelly[4]
Karen Schlangen[5]
Steve Gomez[4]

Naval Health Research Center
P. O. Box 85122
San Diego, CA 92186-5122

Walter Reed Army Institute of Research
Washington, DC 20307-5100

[1] **Report No. 95-6, supported by the Office of Naval Technology and the Naval Medical Research and Development Command, Department of the Navy, under work units 63706N M0096.002-6201 and 62233N MM33P30.002-6005. The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of the Navy, Department of the Army, Department of Defense, or the U.S. Government. Approved for public release, distribution unlimited.**

The authors wish to thank Drs. David Thorne, Diane Williams, Steven Linnville (LT MSC USNR) and Gregory Galbicka for their comments on this paper. Portions of these data were presented at the 1992 Meeting of the American Psychological Association.

[2] *Visiting Scientist at the Naval Health Research Center*
[3] *Walter Reed Army Institute of Research*
[4] *Naval Health Research Center*
[5] *San Diego State University*

Summary

Problem

Many laboratory investigations of the effects of stressors on human performance employ performance assessment batteries consisting of sequential tests of isolated cognitive attributes and abilities. However, operational jobs usually require concurrent performance of two or more tasks, each with its own set of requirements and consequences. Thus, conclusions regarding the operational impact of stressors based on traditional tests may be questioned.

Objective

The objective of this study was to compare performance on a traditional sequential test battery with that on a synthetic work task (SYNWORK1) requiring subjects to work concurrently on several tasks.

Approach

Subjects were tested every three hours during 64 hr of sleep deprivation on a battery of tests including both traditional sequential attribute tests and SYNWORK1.

Results

Performance decrements began to appear during the first night of lost sleep, becoming more severe during the second night. Decrements were greatest in the early morning hours, recovering somewhat during the day. Performance decrements were usually more severe on the sequential tests than on comparable measures from SYNWORK1, and sequential tasks showed greater sensitivity to sleep deprivation (day effect on ANOVA). However, SYNWORK1 measures showed a stronger circadian rhythm (time of day effect on ANOVA) than comparable measures from the sequential tests, and the composite SYNWORK1 score showed a stronger circadian rhythm than any individual task from the sequential tests. For the sequential tests, but not SYNWORK1, performance efficiency (throughput) tended to be more sensitive to sleep deprivation than either speed or accuracy. Subjects reported that SYNWORK1 was more interesting than the sequential tests.

Conclusions

These results confirm long-standing reports in the literature that complex and interesting tasks are less sensitive to disruption by sleep deprivation than boring and uninteresting tasks, thus highlighting the importance of motivation in determining the outcome of laboratory performance tests. The synthetic work approach represents an attempt to bring some of the structural and functional complexities of operational environments into the laboratory, permitting rigorous investigation of the effects of stressors on performance under more realistic conditions than those provided by traditional sequential tests of attributes and abilities. Decisions regarding what type of test is appropriate for a given application require consideration of trade-offs between generality and validity of results and cost inherent in different approaches to performance assessment.

Introduction

Techniques for assessment of human cognitive performance span a broad range, from questionnaires, paper-and-pencil tests, computerized tests of cognitive and motor abilities, to simulators and field exercises. Due to the availability of powerful personal computers, recent years have seen a concentration of interest on the middle of this continuum, computerized testing in the form of performance assessment batteries or PABs (Anger, 1990; Englund, Reeves, Shingledecker, Thorne, Wilson, & Hegge, 1985; Kennedy et. al, 1981, Thorne et. al 1985). These batteries consist of tests that are each designed to measure a limited subset of abilities, in relative isolation from one another (Anger, 1990; Perez, Ramsey, Masline, & Urban, 1987). While performance on PABs is sensitive to a wide range of variables, the relationship between such effects and performance under operational conditions is sometimes difficult to characterize. At best, it can be said that if a given variable produces a decrement on PAB performance, it *may* degrade performance under certain operational conditions.

Work situations outside the laboratory, particularly in operational military settings, usually involve performance of more than one task at a time, with priorities assigned by instructions and contingencies associated with performance of the component tasks. Extensive literature demonstrates that performance differs under dual-task conditions, compared to single-task performance (e.g., Chiles, 1982; Damos, 1989). Thus, determination of effects on dual- or multiple-task performance is a highly desirable step in the evaluation of the performance effects of any stressor that might adversely affect military performance. This study compared effects of 64 h of sleep deprivation on performance of a synthetic work task (Synwork1) requiring subjects to simultaneously perform four concurrent tasks (cf. Alluisi, 1967), with performance on several standard PAB measures.

Method

Subjects

The nine male subjects, enlisted personnel in either the Navy or the Marine Corps, were recruited from a variety of sources. Most had completed high school, but none had completed four years of college. Depending on availability, one to three subjects were run on any given week.

Apparatus

Four IBM/AT-compatible computers (Unisys 386, 20 Mhz, color VGA monitor, 80 Mb hard disk,

Microsoft mouse) were used in this study. The computers were located in a single room, with testing stations separated by low partitions.

Sleep deprivation study

On the day subjects were brought into the laboratory, two training sessions were conducted. In these sessions, detailed instructions concerning each test were provided, as well as trial-by-trial feedback on performance during the tests. The duration of training sessions was the same as described below for the testing sessions. When a subject had difficulties with a particular test, extra practice on that test was permitted. Following the training day, no instructions or trial-by-trial feedback were provided for any of the tests, except as described below for Synwork1. Subjects slept in the laboratory, and were awakened at 0600. A Performance Assessment Battery consisting of eleven tests, including Synwork1, was administered at 0900, and every three hours thereafter for the duration of the study.[6] Approximately 110 minutes were required for completion of the battery. We will present data from Synwork1 and nine of the PAB tests.

PAB tests

Simple reaction time. In this test, a small rectangle was displayed in the center of the personal computer (PC) screen. Subjects were instructed to press either button of the mouse and hold it down when the square turned light blue, and to release the button as rapidly as possible when the square turned green. In reaction time terminology, the blue light was the "foreperiod", and the green light the "trigger." Intertrial intervals (ITIs) were 2 s. The box remained blue for a random time ranging from 0.6 s to 1.2 s. Premature releases or presses initiated a new ITI. Sessions terminated after 5 min or 50 trials.

Go-no-go reaction time. This test was similar to the simple reaction time test previously described, except that on 50% of the trials, the square turned red rather than green, remaining red for 1 s at which time it turned blue again. On these trials, the subject was instructed to hold the mouse button until the square returned to blue. Sessions terminated after 5 min or 50 trials.

[6] Test order was as follows: Matrix test (spatial memory; Two-column Addition; Logical Reasoning (3-letter); Simple Reaction Time; Complex Reaction Time; 10-min Break; Four-choice Reaction Time; Digit-symbol substitution; Logical Reasoning (2,3,4-letter); Word Memory; Synwork1; Tapping (sleepiness test).

Four-choice reaction time. This test adhered to the standards described in Perez et al. (1987). A 5.0-inch square divided into four quadrants was displayed in the center of the PC screen. On each trial, a .75 inch plus (+) sign was displayed in the center of a randomly selected quadrant of the square, and subjects were instructed to press the 7, 9, 1, or 3 key of the numeric keypad of the keyboard with the index finger of the preferred hand. The correct key depended on the quadrant in which the plus sign was displayed (e.g., the 7 key for upper left quadrant, 9 for upper right). The plus sign disappeared as soon as the correct key was pressed, reappearing after a 1-s ITI. Test duration was 11 min.

Two-column addition. Five randomly selected two-digit numbers were displayed one above the other in a column in the center of the screen with a line drawn beneath. Characters were double the normal height and width of PC characters. The subject's task was to add the numbers and enter the answer using the number keys on the PC keyboard. When the first digit of the answer was entered, the problem disappeared. As soon as the "Enter" key was pressed, the screen was cleared and a new problem appeared. Test duration was 10 min.

Word memory. In this task, a list of 10 words was presented for 10 s. After a brief delay, 20 probe words were presented, including the 10 words from the list, and 10 nonlist words. Presentation order was random. For each word, the subject was required to press "T" or "U" on the keyboard. Test duration was 10 min or until 10 lists were completed.

Three-letter logic. The subject was presented with a three-letter sequence containing a random permutation of the letters A, B, and C, and two statements about the sequence. For example, one trial might consist of the following screen:

A FOLLOWS B

A DOES NOT FOLLOW C

A C B

The subject's task was to press "T" if both statements were true, and "U" if one or both statements were untrue. A new problem appeared as soon as a key was pressed. Test duration was 20 min.

Two- Three- Four-letter logic. In this task, one third of the trials were identical to the three-letter logic task. On one third of the trials only two letters and one statement was presented, and on the remaining third of the trials, four letters and three statements were presented. Test duration was 5 min or 30 problems.

Matrix. On each trial, the subject was presented with a screen containing 14 asterisks arranged in a 14 by 14 matrix such that each asterisk was in a unique row and column. The display remained on the screen for 1.5 s, then the screen went blank. After 3.5 s, the matrix was re-displayed, either the same as before, or with four of the asterisks moved to new locations. The subject was instructed to press "S" if the matrix was the same, or "D" if it was different. Test duration was 10 min.

Digit-Symbol Substitution. In each session, the digits 0 to 9 were displayed in a row at the top of the screen, and a row of symbols was displayed, one below each digit. The position of the symbols was randomized each session. In each trial, a random digit was displayed and the subject's task was to press the symbol key associated with that digit. A new digit appeared as soon as a key was pressed. Test duration was 5 min.

Tapping. In this task, subjects were instructed to press a mouse button once per second for the duration of the task. To serve as a pacing stimulus, for the first 10 s, a .75-inch (1.9 cm) diameter red spot flashed in the center of the screen for the first 0.2 s of each 1-s period. After 10 s had passed, the screen remained a uniform dark gray for the remainder of the session. Test duration was 5 min.

Synthetic work test: SYNWORK1

The computer program, SYNWORK1 (referred to in the remainder of the paper as Synwork) has been described in detail elsewhere (Elsmore, 1991, 1994; Elsmore, Naitoh, & Linnville, 1992), and will be briefly described here. The program presents four simultaneous tasks on the screen of an IBM-compatible PC. The tasks comprising Synwork were selected to provide a generic work environment in which the operator is required to remember and classify items on demand, perform a self-paced task (arithmetic problems), and monitor and react to both visual and auditory stimuli. No attempt was made to simulate any particular job or system, although several users have suggested that the program provides a reasonable part-simulation of various watchstanding jobs.

Subjects interact with Synwork using a standard mouse, permitting the subject to concentrate on the information on the screen, eliminating the distractions of a keyboard. Synwork allows researchers broad latitude in the choice of task parameters (e.g., subtask difficulty, payoff matrix). The four subtasks are each displayed in one quadrant ("window") of the screen. The program uses a Sound Blaster sound card for presentation of all auditory stimuli. Subjects wore headphones during the Synwork portion of the test session.[7] Correct responses produced a high-pitched "squeaking" sound, and errors produced a low-pitched "burping" sound. The following paragraphs describe the details of each subtask in the present experiment:

Upper left window, Sternberg memory task. Each session, a list of letters (the positive list) was randomly chosen from the alphabet with the exceptions of the letters C, D, M, Q, and V which were not used. Letters were displayed in uppercase in a box at the top of the window. In the present study, the list was 6 letters long. At the same time, a "negative list" of the same length was selected from the remaining letters of the alphabet. At the beginning of the session, the positive list was displayed for only 5 s, after which it was replaced by the words **"RETRIEVE LIST"**. When this message was displayed, clicking the mouse on the list box resulted in display of the list for another 5 sec. The list could be retrieved as many times as required during the session, although a 10-point penalty was charged when the list was retrieved . At the beginning of each trial, a sample letter was displayed in a box in the center of the window. The letter was randomly selected from either the positive or negative list ($p = .5$) The subject's task was to indicate, by clicking the mouse on either the "**YES**" or the "**NO**" box at the bottom of the window, whether the letter was a member of the positive set or not. Ten points were awarded for each correct choice, and a 10-point penalty was imposed for each error. Following each choice the sample letter disappeared for the remainder of the trial. Trials were 20 s in duration.

Upper right window, arithmetic task. A three-column addition task presented two randomly selected three-digit numbers between 100 and 999, with "**0000**" in the answer space. The subject's task was to adjust the answer by clicking on "+" and "-" boxes below each character of the answer. Clicking on a box labeled "**DONE**" at the bottom of the window resulted in the presentation of a new problem, award of 10 points for correct answers, and subtraction of 10 points for errors. There were no time limits for completion of this task or penalties for noncompletion, thus it was completely self-paced.

[7] The Synwork1 program and a detailed reference manual are available on request from the senior author.

Lower left window, visual monitoring. A pointer moved from the center of a graduated scale, 201 pixels in length, towards either end at a fixed rate of 200 msec per pixel. Clicking the mouse on a box labeled "**RESET**" at the top of the window reset the pointer to the center. The subject's task was to prevent the pointer from reaching the end of the scale. Points were awarded for each reset, with the number of points being proportional to the distance of the pointer from the center at the time of reset up to a maximum of 10 points for resets near the ends of the scale. Ten points were deducted for each second the pointer was at either end of the scale.

Lower right window, auditory monitoring. A brief tone was sounded every 5 s. The tone was either of two frequencies, low (1046 Hz) or high (1319 Hz). The subject's task was to click the mouse in a box at the top of the window labeled "**HIGH TONE REPORT**" following a high tone. The probability of high tones was 0.2. Correct responses were those following high tones prior to the next scheduled tone. All other responses were incorrect. Ten points were awarded for each correct response and were deducted for each error.

Results

Because of the differences between tasks, no direct statistical comparisons were made among different measures. Rather, each measure was analyzed separately. Comparisons among measures will only be made with reference to the presence or absence of significant sleep deprivation or time of day effects relative to the baseline for that measure. Table 1 presents the results of repeated-measures ANOVAs for measures of response speed, accuracy, and performance efficiency for all of the PAB tests except the tapping test, and for Synwork subtests from midnight on Day 2 to the end of the experiment (the 2100 session on Day 3). Thus, there were 2 levels for day (Days 2 and 3), and 8 levels for time of day (the 8 sessions conducted each day).

In Table 1, the Response Speed measure used was the reciprocal of the average time required to make a correct response. Response Accuracy was simply the percent correct responses on each task. As an estimate of Performance Efficiency for the PAB tests, *throughput*, defined by Thorne (1983) as correct responses per working minute, was used. For Synwork, the composite score and the scores for each individual task were used. This table does *not* show any clear differences between the PAB tests and

Synwork. It is interesting to note that performance efficiency seemed to be a more sensitive indicator of sleep deprivation effects than either of the two measures upon which it is based.

Table 1
Analysis of Variance p-Values for Day and Time of Day (Circadian Rhythm)[8]

	Response Speed		Response Accuracy		Performance Efficiency	
Test	Day	Time of Day	Day	Time of Day	Day	Time of Day
Matrix	•	•	**0.030**	**0.026**	•	•
Two-column Addition	0.069	0.160	**0.006**	0.332	**0.001**	**0.017**
Four-choice	**0.029**	0.199	0.319	0.373	**0.004**	**0.027**
Simple RT	0.974	0.434	*	*	**0.036**	0.196
Go-No-Go RT	0.080	0.060	0.074	0.227	**0.011**	0.080
Logic (3)	**0.010**	0.296	0.084	0.165	**0.001**	0.129
Logic (2,3,4)	0.195	0.337	0.056	0.135	0.604	0.370
Digit-Symbol Substitution	0.115	0.371	0.109	0.483	**0.013**	0.197
Word Memory	0.292	0.386	**0.002**	0.134	0.385	0.093
Synwork overall rate	0.211	**0.028**	*	*	*	*
Synwork comp. score	*	*	*	*	**0.019**	**0.013**
Synwork memory	0.133	0.367	0.408	**0.046**	0.279	**0.041**
Synwork addition	0.077	**0.003**	**0.013**	0.145	**0.039**	**0.015**
Synwork visual mon.	**0.008**	0.384	*	*	**0.017**	0.141
Synwork auditory mon.	0.124	0.151	0.096	0.093	0.131	0.142

Performance on most tasks reached a peak at 2100 on Day 1. At this point, the tasks were well-learned, and sleep deprivation was minimal. Beginning with the next session, sleep loss began to take is toll, and performance deteriorated. To characterize these performance decrements, a series of t-tests was conducted in which performance at each test time on Days 2 and 3 was compared with the 2100 session on Day 1. Thus, significant effects represent departure from the optimal performance for that task.

[8] Degrees of freedom: Day (1, 8); Time of Day (7,56), adjusted with the Greenhouse-Geiser correction for repeated measures.
• Missing data * Not defined

Table 2
Percentage of Baseline[9]

Response Speed

Time of day →	Day 2 0000	0300	0600	0900	1200	1500	1800	2100	Day 3 0000	0300	0600	0900	1200	1500	1800	2100
Two-column Addition	104.9	87.5^{++}	78.0^{++}	72.7^{++}	74.7^{+}	82.5^{++}	83.4^{++}	79.7^{++}	88.3	74.4^{+}	76.4^{+}	81.9	74.5^{++}	75.2^{++}	90.3	82.3^{++}
Four-choice	92.2	86.8^{++}	60.5^{++}	74.0^{++}	66.1^{++}	77.6^{++}	88.4	70.5	60.4^{++}	48.8^{++}	48.7^{++}	57.3^{++}	63.6^{++}	70.0^{++}	64.0^{++}	78.4^{++}
Simple RT	96.1	77.6^{+}	54.5^{++}	77.3^{++}	74.6^{++}	71.8^{++}	74.3^{++}	77.7^{+}	56.6^{+}	65.7^{+}	56.8^{+}	65.1^{+}	69.0^{+}	68.7^{+}	72.5^{+}	67.6^{+}
Go-No-Go RT	93.4	84.1^{++}	69.3^{++}	79.0^{++}	74.6^{++}	72.7^{+}	80.3^{++}	82.0^{++}	68.7^{++}	74.0^{+}	69.9^{++}	72.0^{++}	76.7^{+}	63.3^{++}	74.0^{++}	75.5^{++}
Logic (3)	80.1^{+}	81.4	69.4	117.4	45.8^{++}	54.5^{++}	48.4^{++}	54.5^{++}	42.9^{++}	92.9	42.3^{++}	47.2^{++}	49.3^{++}	49.8^{++}	52.4^{++}	66.9^{+}
Logic (2,3,4)	70.6^{++}	68.4^{++}	71.8^{+}	61.0^{++}	50.6^{++}	50.1^{++}	60.9^{++}	60.1^{++}	58.6^{++}	73.0	116.8	77.2	63.4^{++}	51.0^{++}	76.6	98.5
Digit-Symbol Substitution	96.1	98.8	71.8^{++}	65.8^{++}	71.5^{+}	67.0^{+}	79.4	75.5	67.4^{+}	47.6^{++}	75.9	51.5^{++}	57.5^{++}	55.5^{++}	56.3^{+}	63.6^{++}
Word Memory	98.6	94.7^{+}	94.5	59.0^{+}	84.3^{+}	69.9^{++}	96.3	89.4^{+}	70.0^{+}	101.8	93.4	68.3^{+}	80.7	89.1	158.5	109.0
SYN1 overall rate	102.2	94.2	82.5	88.7	102.1	98.6	102.3	91.2	85.1^{++}	82.3^{+}	79.6^{++}	85.9	101.2	99.1	103.8	94.7
SYN1 memory	103.1	105.7^{+}	87.0^{+}	94.1	98.5	96.7^{+}	99.2	96.8	76.6^{+}	87.3^{+}	71.0^{+}	123.6^{+}	88.5	90.4	119.0	123.5
SYN1 addition	104.3	97.7	78.3^{++}	92.2	108.7	107.9	110.3	88.1	80.6^{++}	74.1	70.0^{++}	83.7^{+}	100.3	102.7	112.5	98.2
SYN1 visual mon.	109.6	110.8	111.4	107.6	108.1	108.0	117.5	143.8	145.7	159.5	142.8	138.0	124.6	138.4^{+}	125.2^{+}	129.0^{+}
SYN1 auditory mon.	99.4	92.2^{++}	89.2^{++}	97.0	96.5^{++}	92.0^{++}	104.4	93.8^{+}	96.3^{+}	88.8^{+}	89.3^{++}	89.7^{+}	94.5^{+}	95.4	96.9	90.0^{++}

Response Accuracy

Time of day →	Day 2 0000	0300	0600	0900	1200	1500	1800	2100	Day 3 0000	0300	0600	0900	1200	1500	1800	2100
Two-column Addition	101.8	93.9	93.1	85.6^{+}	97.2	87.7	91.5	90.4^{++}	76.8^{+}	82.0	72.5	79.5^{+}	89.3	79.3^{+}	92.3	80.1^{+}
Four-choice	99.9	99.8	96.0^{+}	98.9	98.7	96.8	95.2	99.2	97.5	94.6	95.7^{+}	95.4^{+}	97.6^{+}	97.0^{+}	97.2	98.8
Go-No-Go RT	99.7	98.1	93.5	92.5^{+}	93.7	97.9	98.0	99.7	92.3	96.8	94.1	93.2^{+}	90.2^{++}	92.1^{++}	97.4	95.1
Logic (3)	98.4	91.7	91.9	81.7^{+}	96.7	94.4^{+}	97.8	94.7	93.1^{+}	88.0	74.6^{+}	87.5^{+}	88.1^{+}	85.9^{+}	86.7^{+}	84.9^{+}
Logic (2,3,4)	97.4	96.4^{+}	89.5^{+}	97.0	96.0	92.9	101.3	98.2	87.0	88.2	76.8^{++}	82.8^{+}	90.8	91.1	83.7^{+}	91.0
Digit-Symbol Substitution	99.8	98.7	96.1	96.2	93.1	99.2	99.9	97.4	98.1	90.8	86.8	97.8	94.9	96.8	95.7^{++}	98.6
Word Memory	93.1^{++}	86.9^{++}	79.0^{++}	81.9^{++}	85.4^{++}	84.5^{++}	93.8	88.9^{++}	75.2^{++}	69.6^{++}	74.1^{++}	76.5^{++}	80.4^{++}	80.3^{++}	81.2^{++}	88.1^{++}
Matrix	102.7	100.0	90.2^{++}	88.9^{++}	89.2^{+}	92.3	93.7	94.4	90.8^{+}	89.2^{++}	86.4^{+}	80.4^{++}	92.9	88.1^{+}	96.1	93.6
SYN1 memory	100.9	102.7	91.5	100.4	98.3	103.3	110.5	105.0	99.4	113.4	91.0	101.1	122.0^{++}	112.4^{+}	105.5	84.9
SYN1 addition	99.7	93.9	90.3^{+}	91.1^{+}	97.4	92.7	97.9	88.5	87.9^{+}	78.6	75.9^{++}	87.4^{+}	93.2^{++}	89.5^{++}	96.1	92.2^{+}
SYN1 auditory mon.	101.8	100.2	98.5	98.8	101.5	99.5	100.6	98.0	97.0	97.1	89.4	99.1	97.4	100.5	100.7	99.0

[9] Percentage difference from baseline (see text). $^{+}$, $p < .05$; $^{++}$, $p < .01$

Table 2 (page 2)

Performance Efficiency

	Day 2								Day 3							
Time of day →	0000	0300	0600	0900	1200	1500	1800	2100	0000	0300	0600	0900	1200	1500	1800	2100
Two-column Addition	105.1^{+}	82.6^{+}	64.7^{++}	62.1^{++}	67.8^{+}	70.7^{++}	75.9^{++}	71.7^{++}	69.3^{++}	57.2^{++}	48.3^{++}	58.0^{++}	61.8^{++}	60.8^{++}	81.4^{++}	66.1^{++}
Four-choice	91.8	86.5^{++}	58.1^{++}	72.8^{++}	65.1^{++}	75.2^{++}	82.7^{++}	69.9^{+}	58.8^{++}	46.1^{++}	46.7^{++}	55.2^{++}	61.9^{++}	68.3^{++}	62.6^{++}	77.0^{++}
Simple RT	94.3	77.0^{+}	47.7^{++}	76.7^{++}	73.4^{++}	71.2^{++}	73.1^{++}	76.6^{+}	49.7^{++}	64.8^{+}	55.9^{++}	63.9^{++}	67.6^{+}	67.5^{+}	71.2^{+}	66.6^{++}
Go-No-Go RT	92.8	82.1^{++}	65.1^{++}	72.3^{++}	69.7^{++}	70.5^{+}	78.3^{+}	80.6^{++}	62.5^{++}	70.5^{++}	64.7^{++}	66.1^{++}	67.8^{++}	57.8^{++}	71.4^{++}	70.7^{++}
Logic (3)	78.0^{+}	70.1^{++}	58.7^{++}	72.9	43.7^{++}	50.8^{++}	46.6^{++}	50.2^{++}	39.6^{++}	59.8	29.2^{++}	40.5^{++}	43.0^{++}	40.2^{++}	44.5^{++}	52.3^{++}
Logic (2,3,4)	67.7^{++}	64.5^{++}	59.7^{++}	57.8^{++}	47.7^{++}	45.6^{++}	60.2^{++}	57.7^{++}	45.8^{++}	50.5^{+}	73.6^{++}	59.9^{++}	54.3^{++}	44.9^{++}	57.8^{++}	78.6
Digit-Symbol	98.6	102.8	70.4^{++}	68.9^{+}	64.5^{+}	70.0^{+}	82.7	79.3	71.1^{+}	50.4^{++}	47.5^{++}	54.3^{++}	58.2^{++}	57.5^{++}	55.9^{++}	66.4^{++}
Word Memory	91.4^{++}	82.6^{++}	71.9^{++}	62.7^{++}	75.1^{++}	76.5^{+}	100.1	90.1	65.0^{++}	65.7^{++}	65.5^{++}	63.9^{++}	70.2^{++}	71.1^{++}	99.4	92.2^{+}
SYN1 composite score	104.4	103.0	84.8	88.3	106.9	102.6	107.6	92.7	83.5^{+}	84.4	59.4^{++}	78.3	104.4	96.7	109.7	91.9
SYN1 memory	114.2	125.6	80.9	109.1	103.7	126.7	153.9	132.9	119.7^{++}	158.3^{+}	75.0^{++}	101.3	191.6	153.1^{+}	129.7	50.3
SYN1 addition	103.4	84.5	58.6^{++}	76.5^{+}	101.8	91.3	105.7	73.3	67.9	51.4	37.6	74.6	86.4	81.0	103.6	83.0
SYN1 visual mon.	101.6	102.4	100.5	91.9	106.1	104.5	104.6	102.0	94.2	89.8	67.4	78.8	96.6	105.3	107.6	99.7
SYN1 auditory mon.	111.0^{++}	119.8^{++}	124.2^{+}	105.1	122.0^{++}	109.7^{++}	98.6	86.7	102.2	100.3	73.2^{+}	93.7	115.1^{+}	97.3	107.6^{+}	99.6

Table 2 shows normalized group means for response speed, accuracy, and performance efficiency for all of the tasks except the tapping task. These values were calculated by converting the scores for each subject and test to the percentage of baseline for that subject and test, where baseline was the mean score for the 5 sessions on Day 1, then averaging these normalized values to obtain a group mean. To provide statistical estimates of the significance of departures from baseline, t-tests were conducted for each test comparing performance at each session on Days 2 and 3 with the last session on Day 1. In almost all cases, significant departures from baseline represented decreases. In a few cases, however, most notably SYNWORK visual monitoring response speed, there were significant increases. To simplify comparisons among measures, the differences are summarized in Table 3, which shows the percentages of sessions deviating from baseline.

Table 3

Percentage of Sessions Significantly Different From Baseline

	Day 2	Day 3
	Response Speed	
PAB	75.0	75.0
Synwork	22.5	47.5
	Response Accuracy	
PAB	28.1	54.7
Synwork	8.3	33.3
	Performance Efficiency	
PAB	84.4	96.9
Synwork	17.5	22.5

For the purposes of the present paper, the most important fact to be seen in Tables 2 and 3 is that Synwork measures showed fewer significant departures from baseline than the PAB measures. This was true regardless of response measure or day. Secondly, as expected, greater effects were seen for most measures on Day 3 than on Day 2. Third, some apparent time-of-day effects were evident, mostly on

Day 3. Finally, Table 2 suggests that accuracy measures were the least sensitive to sleep deprivation, although performance efficiency for Synwork was also relatively insensitive.

A number of figures will better illustrate some of the differences between PAB and Synwork performance. Figure 1 shows normalized group mean response speed for Synwork addition and the PAB two-column addition tasks. For both tasks, response speed was calculated as the reciprocal of the time required to complete a correct response. Sleep deprivation clearly had a much more dramatic effect on the PAB task than on the Synwork task. Both declined during the morning hours, but only the Synwork task recovered during afternoon and evening sessions.

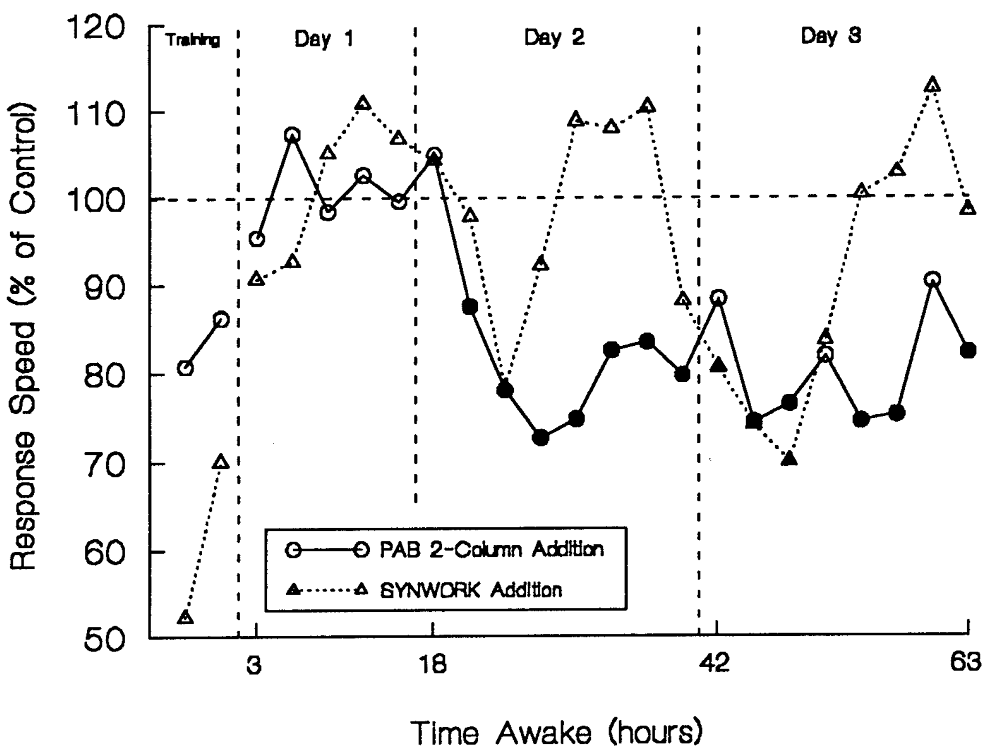

Figure 1. Normalized plots of response speed (reciprocal of the time required to complete a correct response). Points are means of nine subjects. Mean scores from Day 1 were used as control values. Filled points are significantly ($p < .05$) different from the last session on Day 1. Vertical reference lines indicate midnight.

In Figure 2, normalized accuracy scores on the Synwork addition subtask are compared with accuracy on the two-column addition PAB task. This figure shows that accuracy on both tasks declined as sleep deprivation increased, reaching a minimum during the early morning hours of Day 3. Unlike the most other comparisons made in this paper, there was no clear difference between the PAB measure and the Synwork measure, although there was a tendency for the Synwork measure to recover more than the PAB-based measure in the afternoon and evening of each day.

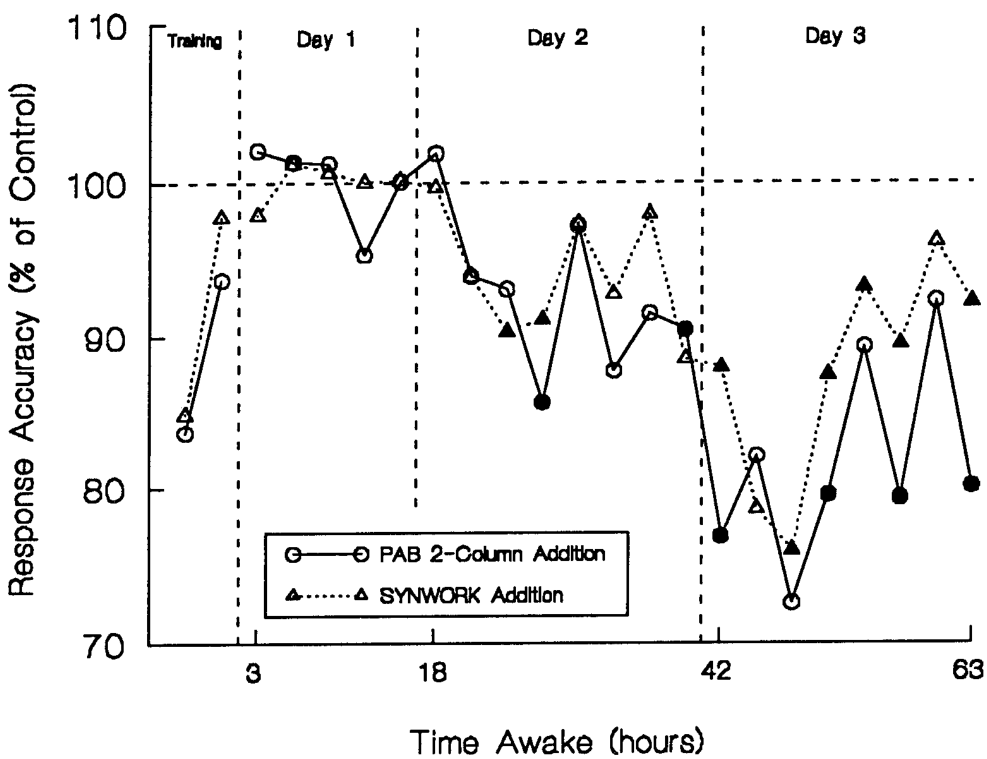

Figure 2. Normalized accuracy scores of the Synwork addition task and 2-column addition. Points are means of nine subjects. Filled points are significantly ($p < .05$) different from the last session on Day 1.

As indicated earlier, measures of performance efficiency appear to be most sensitive to sleep-deprivation effects. To illustrate this point, as well as the difference between PAB and Synwork

performance, Figure 3 compares average throughput of 8 PAB tasks with the Synwork composite score (i.e., the sum of the 4 subtask scores). This figure clearly shows that PAB throughput decreased as sleep deprivation increased, remaining substantially below baseline levels until the end of the experiment. Synwork performance, on the other hand, decreased in the morning hours and returned to baseline in the afternoon and evening.

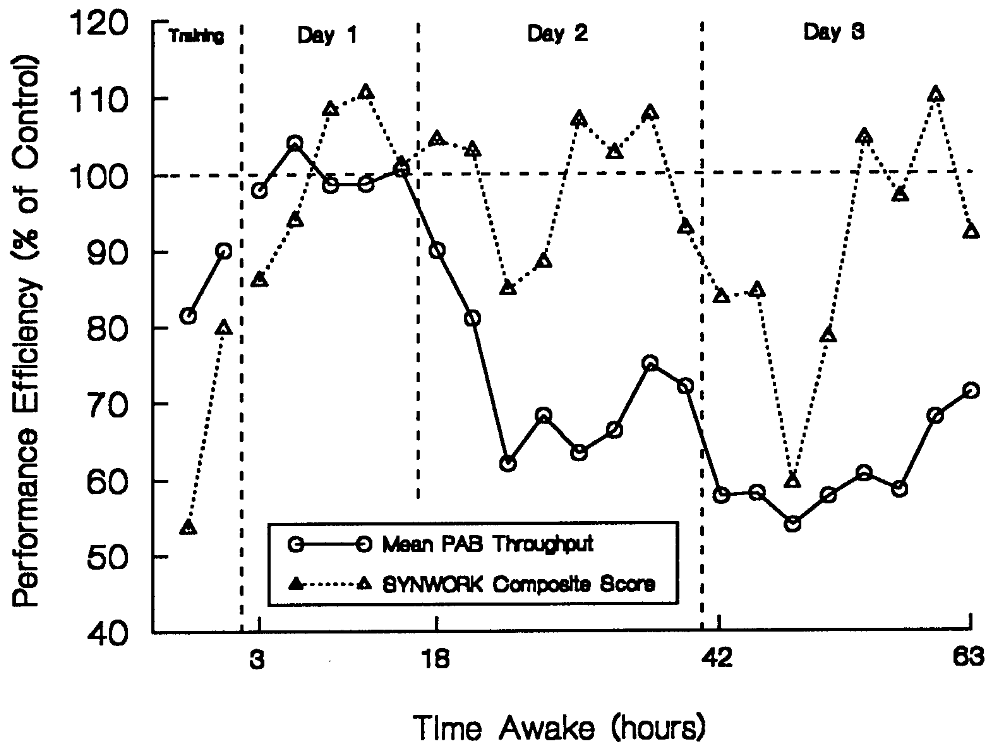

Figure 3. Normalized performance efficiency scores. PAB scores were averaged across eight different PAB tasks. Points are means of nine subjects.

The measures discussed up to this point focus on *task performance*, although failures to respond do contribute to response speed and performance efficiency measures. In real-world situations, failures to respond to a situation may prove disastrous. Thus, it is appropriate to focus directly on this type of error, that is to say "lapses." Two quite different measures of failure to respond, lapses in the tapping task

(i.e., inter-response intervals > 5 s) and pointer-reset failures on the visual monitoring task in Synwork, are plotted in Figure 4.

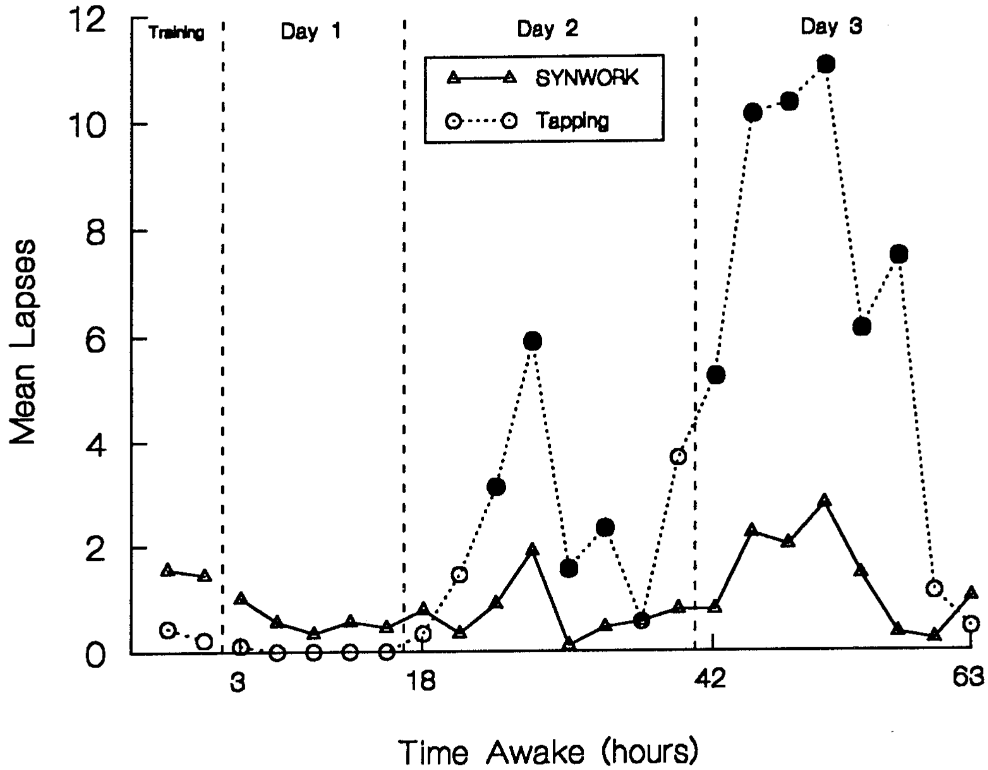

Figure 4. Response omissions (Lapses). Pointer-reset failures on the Synwork visual monitoring task and lapses (interresponse times > 5 sec) on the tapping task. Points are means of 9 subjects.

Both measures follow the same pattern, showing large and progressively greater increases on Days 2 and 3. However, the increases were considerably larger for the tapping task. As was true for the previous data, these measures were affected following the first night of sleep deprivation, with much larger effects following the second night of lost sleep. Statistically, there were no significant effects for Synwork visual monitoring lapses, but both day ($\underline{F}(1,8 = 27.7< .01)$ and time of day ($\underline{F}(3.5,28) = 5.31< .01$) were

significant for the tapping task. Other response-omission measures for both PAB and Synwork showed similar effects, with the PAB tests generally showing larger effects of sleep deprivation.

Discussion

Variables controlling test performance.

The data presented here demonstrate clear differences between sleep-deprivation effects for tests presented in the traditional sequential performance assessment battery format and the concurrent format provided by Synwork. These results are consistent with the early work of Wilkinson (1964) who stated in an article summarizing a large number of sleep deprivation experiments:

> A task will be vulnerable to sleep deprivation (1) as it is complex, and (2) as it is lacking in interest. Of the two factors, that of incentive may be the more influential, such that a highly complex task may be little affected by sleep deprivation if it is complex in an interesting and rewarding way. (p. 175)

The relationships between task complexity, motivation, and sleep deprivation were also discussed by Fröberg (1985) who concluded that both task characteristics and motivational factors interact to determine the impact of sleep loss on performance. These observations are contrary to the simplistic view that more complex tasks are necessarily more sensitive to disruption by sleep deprivation (or other stressors). Indeed, the present data support the view that a dissociation exists between the complexity of a task and its resistance to disruption by sleep deprivation (and probably other stressors as well), and that one must turn to motivational factors, as suggested 30 years ago by Wilkinson, for an explanation of the relative insensitivity of Synwork to disruption by sleep deprivation and circadian rhythms.

One of the most potent motivational variables in Wilkinson's (1964) experiments was immediate knowledge of results. Regardless of the complexity of a task, subjects who received information on their performance immediately following each test were less affected by sleep loss than those who received no information about their performance. In typical PAB tests, subjects are *not* given knowledge of results, which undoubtedly contributes to the relative fragility of performance on these tests. In Synwork, on the other hand, subjects are fully informed of their progress *during* the test. From an operational point of view, these findings suggest that tasks having relatively immediate consequences might be expected to fare better under stressful conditions than those with less immediate effects.

Synwork is frequently treated by subjects as though it were a game. Indeed, it has many of the attributes of a game, including simultaneous action in several locations, motion (though it is very simple), the necessity of developing a strategy for "playing the game," awarding of points for performance, and competition (with one's self or others). Thus, Synwork provides several potential sources of reinforcement. Points can serve as generalized reinforcers based on their association with other reinforcers such as praise and competition. The simple opportunity to choose courses of action has been demonstrated to be reinforcing (e.g., Catania, 1975). The relatively complex visual and auditory environment provided by Synwork may also be a source of reinforcement. This aspect of Synwork, and other game-like testing procedures, may have practical consequences for the conduct of research studies requiring repeated testing for extended periods. Subjects are motivated (or at least less reluctant) to return to the testing situation, resulting in more reliable data collection.

One place where Synwork proved comparable to a standard PAB measure was in addition accuracy, which degraded in a highly similar manner for both Synwork and the two-column addition task. While, based on these data, it may be tempting to conclude that similar processes were in effect in both tests, such a conclusion is premature. It is interesting to note that the addition subtask of Synwork is one aspect of the task that is not temporally constrained. Thus, even *within* a complex and generally motivating task, those aspects of the task that are not tightly controlled may be vulnerable to disruption.

Circadian Rhythm (time-of-day) effects.

While there were some clear instances of circadian rhythmicity in the present study, there were also many measures which showed little reliable variation with time of day (Table 1). Inspection of Table 2 and Figure 3 suggests that the lack of circadian effects for most of the PAB tests was that PAB performance tended to be suppressed for most of Days 2 and 3. On the other hand, when Synwork performance failed to be rhythmic it was because there was little change from baseline. Thus a given performance can fail to demonstrate a circadian rhythm in two ways, illustrated by curves A and C in Figure 5. Curve B, showing strong circadian rhythmicity, is intermediate between these extremes. Elsmore and Conrad (1979, reported in Elsmore and Hursh, 1982) presented some data suggesting that the three curves shown in Figure 5 may lie on a continuum determined by the amount of reinforcement maintaining a performance. In this experiment, rhesus monkeys earned food by pressing on a panel on the side of their cage. During test sessions, the panel could be illuminated one of four colors, each associated with a different frequency of food presentation, and test sessions were conducted six times each day, equally spaced around the clock.

Under these circumstances, panel pressing maintained by a high frequency of reinforcement showed little variation with time of day, while panel pressing maintained by a low frequency of pellet presentation was strongly rhythmic. While it is a considerable leap from a study with nonhuman primates to the present study, it is not unreasonable to speculate that similar motivational processes may have been operating to determine circadian rhythmicity, particularly in work rate. That is, Synwork performance generally showed less departure from baseline than the PAB tasks (see Figures 1 and 3), which would be predicted from subjects' reports that they were more strongly motivated to work on Synwork.

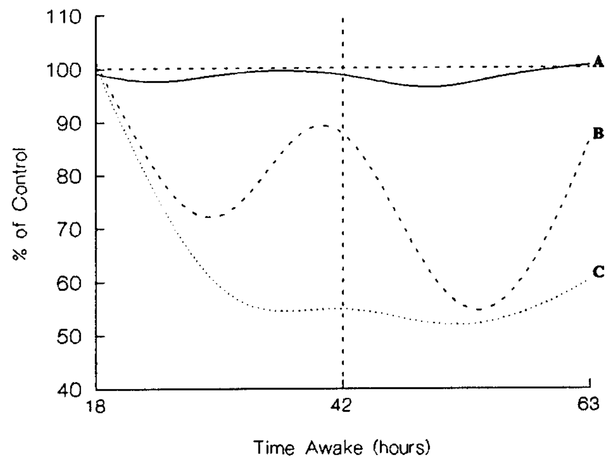

Figure 5. Sleep deprivation effects on three hypothetical performances.

Prediction of operational performance: The Performance Assessment Hierarchy.

The present data demonstrate that the structure of a test is critical in determining the degree to which the test is sensitive to sleep deprivation, and by extension, other stressors. Which then are the "best" tests to use? If this question is not qualified in some manner, there may be no reasonable answer. Figure 6 presents a schematic representation of the trade-offs that come into play as one traverses the continuum of performance assessment techniques from performance assessment batteries to field tests. Sensitivity and generality of tests are greatest at the base of the triangle. Tests at this level evaluate behavioral functions that are common to many different types of operational performance. As one approaches the apex of the triangle, test situations approach operational situations of interest. Interpretability or "validity" of test data increase, however, so do the costs (money, personnel, and time) of conducting the test. "Turf" issues, (e.g., conflicts between training and research missions), also come into play when researchers desire access to system simulators or field exercises. As one moves from the controlled environment of the laboratory into the messy operational world, control of relevant variables decreases, resulting in a concomitant decrease in ability to detect significant effects of stressors.

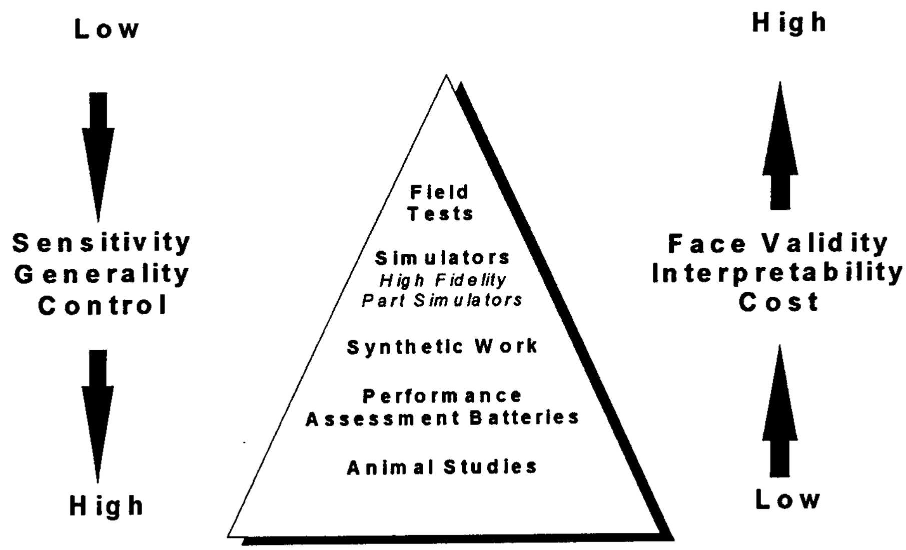

Figure 6. Performance assessment hierarchy and tradeoffs.

Thus, the question of what type of test is best in any given situation must take these trade-offs into account. The manner in which a research question is framed is central to the determination of what type of test is appropriate. For example, the question, "How does sleep deprivation affect performance?" does not specify what kind of performance, and thus calls for tests of broad generality, that is, PAB-type tests. On the other hand, the question, "How does sleep deprivation affect performance *of sonar operators*?" requires a test with structural and functional properties that more closely approximate the operational situation faced by sonar operators. To the degree the testing situation differs from the operational setting, detection of a performance decrement on a test merely *suggests* that a similar decrement might exist under operational conditions, and it establishes a requirement for further research under more realistic conditions. For a multitude of reasons, in many cases it may be impossible or impractical to take this next step.

The synthetic work approach represents an attempt to begin the process of introducing the structural and functional characteristics of operational jobs into laboratory performance testing. Synwork was not intended to simulate any particular operational job, but rather to be a proof of concept. It is theorized, and these data suggest, that the synthetic work approach may provide a "happy medium" between standard computer-based performance assessment batteries and part-task or full simulations and field tests. The approach retains the economy and convenience of PC-based testing, while increasing the

likelihood that test data will be relevant to the operational world. It remains for future work to develop and refine the approach, through the development of test systems targeted to particular operational environments.

References

Alluisi, E. A. (1967). Methodology in the use of synthetic tasks to assess complex performance. *Human Factors,* **9**, 375-384.

Anger, W. K. (1990). Worksite behavioral research: Results, sensitive methods, test batteries, and the transition from laboratory data to human health. *Neurotoxicology,* **11**, 629-720.

Catania, A. C.. (1975). Freedom and knowledge: An experimental analysis of preference in pigeons, *Journal of the Experimental Analysis of Behavior*, **24**, 89-106.

Chiles, W. D. (1982). Workload, task, and situational factors as modifiers of human performance. In E. A. Alluisi & E. A. Fleischman, (Eds.), *Human Performance and Productivity: Stress and Performance Effectiveness* (pp. 11-29). Hillsdale, NJ: Lawrence Earlbaum Associates.

Damos, D. L. (1989). The effects of high information processing loads on human performance. *Proceedings of the 7th Aerospace Behavioral Technology Conference* (pp. 51-54), Warrendale PA: SAE.

Elsmore, T. F. (1991). A synthetic work environment for PC-compatible microcomputers. *Proceedings of the 1991 Medical Chemical Defense Bioscience Review* (pp. 351-354). Aberdeen Proving Ground, MD: U.S. Army Medical Research Institute of Chemical Defense.

Elsmore, T. F. (1994). Synwork1: A PC-based tool for assessment of performance in a simulated work environment. *Behavior Research Methods, Instruments, & Computers*, **26**, 421-426.

Elsmore, T. F. & Conrad, D. G. (1979). Rhythms in learned behavior of monkeys: Effects of frequency of reward. *Chronobiologia*, **6**, 95.

Elsmore, T. F. & Hursh, S. R. (1982). Circadian rhythms of animals under laboratory conditions. In F. M. Brown and R. C. Graeber, (Eds.), *Rhythmic Aspects of Behavior*, (pp. 273-310). Hillsdale, NJ: Lawrence Earlbaum Associates

Elsmore, T. F., Naitoh, P., & Linnville, S. (1992). Performance assessment in sustained operations. Naval Health Research Center, Technical Report 92-30, San Diego, CA: Naval Health Research Center.

Englund, C. E., Reeves, D. L. Shingledecker, C. A., Thorne, D. R., Wilson, K. P., & Hegge, F.W. (1985). *The Unified Tri-Service Cognitive Performance Assessment Battery (UTC-PAB): Design and Specification of the Battery.* (JWGD^3 MILPERF Report No. 85-1). Fort Detrick, MD: U. S. Army Medical Research and Development Command.

Fröberg, J. E.. (1985). Sleep deprivation and prolonged working hours. In S. Folkard & T.H. Monk (Eds), *Hours of Work* (pp. 67-75). New York: John Wiley & Sons.

Kennedy, R. S., Bittner, A. C., Carter, R. C., Krause, M., Harbeson, M. M., McCafferty, D. B., Pepper, R. L., & Wilker, S. F. (1981) *Performance evaluation tests for environmental research (PETER): Collected papers*. (NBDL - 80R008). New Orleans, LA: Naval Biodynamics Laboratory.

Perez, W. A., Masline, P. J., Ramsey, E. G., & Urban, K. E. (1987). *Unified Tri-Services Cognitive Performance Assessment Battery: Review and Methodology*. (AAMRL-TR-87-007). Dayton, OH: Armstrong Aerospace Medical Research Laboratory.

Thorne, D., Genser, S., Sing, H., & Hegge, F. (1983). Plumbing human performance limits during 72 hours of high task load. In: Proceedings of the 24th DRG Seminar on Humans as Limiting Elements in Military Systems. Defence and Civil Institute of Environmental Medicine, Canada. Pp 17-40.

Thorne, D., Genser, S., Sing, H., & Hegge, F. (1985). The Walter Reed performance assessment battery. *Neurobehavioral Toxicology and Teratology*, 7, 415-418.

Wilkinson, R. T. (1964). Effects of up to 60 hours of sleep deprivation on different types of work. *Ergonomics*, 7, 175-186.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE FEBRUARY 1995	3. REPORT TYPE AND DATE COVERED Interim 1 OCT 93 - 30 SEP 94

4. TITLE AND SUBTITLE A comparison of the effects of sleep deprivation on synthetic work performance and a conventional performance assessment battery

5. FUNDING NUMBERS
Program Element:
Work Unit Number:
63706N M0096.002-6201
62233N MM33P30.002-6005

6. AUTHOR(S) T.R. Elsmore, F.W. Hegge, P. Naitoh, T. Kelly, K. Schlangen, and S. Gomez

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Health Research Center
P. O. Box 85122
San Diego, CA 92186-5122

8. PERFORMING ORGANIZATION
Report No 95-6

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Naval Medical Research and Development Command
National Naval Medical Center
Building 1, Tower 2
Bethesda, MD 20889-5044

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Research on the effects of stressors on human performance often uses performance assessment batteries (PABs) of sequential tests of isolated cognitive abilities. However, operational jobs usually require concurrent performance of two or more different tasks. This study compared performance during 64 hr of sleep deprivation on a traditional PAB with that on a synthetic work task (SYNWORK1) involving several concurrent tasks. Subjects were tested every 3 hr on a battery including both traditional tests and SYNWORK1. Performance decrements were more severe on the sequential tests than on comparable measures from SYNWORK1. Subjects reported that SYNWORK1 was more interesting than the sequential tests. These results confirm long-standing reports that tasks which are both complex and interesting are less sensitive to sleep deprivation than uninteresting tasks, highlighting the importance of motivation in the outcome of laboratory performance tests. The synthetic work approach attempts to bring some of the complexities of operational environments into the laboratory, permitting investigation of the effects of stressors on performance under more realistic conditions than those provided by traditional PABs. Decisions regarding what type of test is appropriate for a given application require consideration of trade-offs between generality and validity of results and cost inherent in different approaches to performance assessment.

14. SUBJECT TERMS Circadian rhythms, sleep deprivation, cognitive performance, performance assessment battery, synthetic work

15. NUMBER OF PAGES 24

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited